This Page Is Inserted by IFW Operations and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

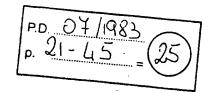
IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents will not correct images, please do not report the images to the Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

XP-002105956

Mol. Biol. Med. (1983) 1, 21-45



Sequence Analysis of the 17,166 Base-pair *EcoRI* fragment C of B95-8 Epstein-Barr Virus

A. T. Bankier, P. L. Deininger, P. J. Farrell and B. G. Barrell

MRC Laboratory of Molecular Biology, MRC Centre Hills Road, Cambridge CB2 2QH, England

(Received 4 March 1983)

Summary In order to provide a framework for understanding the molecular biology of Epstein-Barr virus (EBV), we are determining the DNA sequence of the virus and studying the organization of genes on the viral genome. In this paper we report the DNA sequence of the EcoRI C fragment of the B95-8 strain of EBV. The large (approximately 13.6 kb‡) deletion in this strain has been located by comparison with the DNA sequence of EBV isolated from Raji cells. The sequence has been analysed for possible protein coding regions and transcriptional control sites. At least eight large open reading frames are found, some of them associated with canonical promoter and polyadenylation sequences. The sequences of some of the encoded proteins suggest that they are membrane proteins. It is known that antibodies to major membrane glycoproteins of EBV can neutralize infection in tissue culture. A possible relationship between some of the encoded proteins and the major membrane glycoproteins of the virus is discussed.

Introduction

Relatively little is known about the molecular biology of Epstein-Barr virus (EBV) despite its medical importance as the causative agent of infectious mononucleosis and its association with Burkitt's lymphoma and nasopharyngeal carcinoma (reviewed by Epstein & Achong, 1979; Tooze, 1980). An understanding of the structure of the viral genes and the control of their expression will be extremely useful in developing vaccines or synthetic immunogens against EBV infection. Since EBV transforms some human cells in tissue culture an analysis of the structure and function of the genes required to effect this transformation should make an important contribution to our understanding of natural human carcinogenesis. The virus also exhibits long-term latent infections in humans and provides a system in which to study the importance of this type of infection. The alternative growth patterns of the virus (latent and

[†] Author to whom all correspondence should be addressed.

[‡] Abbreviation used: kb, 103 bases.

productive infections) and their tissue specificity within the human body provide an interesting problem in the control of gene expression since different groups of viral genes are thought to be expressed in the different systems.

Progress in understanding the organization of EBV gene expression has been complicated by both the large size of the viral genome and the lack of a simple permissive cell culture system for lytic growth of the virus. The use of diterpene ester tumour promoters to induce EBV synthesis in transformed cell lines (zur Hausen et al., 1978,1979) has made it much easier to study viral gene expression and the cloning of the viral DNA has made analysis of the viral genome organization possible. Restriction enzyme maps for several enzymes (EcoRI, BamHI, Sall and HindIII) have been established and clone libraries have been made by recombinant DNA techniques for the EcoRI and BamHI restriction enzyme fragments (Dambaugh et al., 1980; Skare & Strominger, 1980; Arrand et al., 1981).

The major features of the genome, which is approximately 170 kbt in length, are two unique regions, the small unique region (U_S) and the large unique region (U_L) , which are separated by a variable (up to 12 copies) number of direct repeats each 3.07 kb in length. Another class of direct repeats each about 540 bases in length, again varying in number, are found at both ends of the linear viral DNA. These ends can circularize to form an intracellular episomal form. An approximate transcription map has been constructed by hybridization of mRNA to the EcoRI and BamHI restriction enzyme fragments (Hummel & Kieff, 1982a). Using hybrid selected translation, the genes for some EBV proteins have been assigned to particular restriction fragments (Hummel & Kieff, 1982b). In addition the genes for two small RNAs transcribed by RNA polymerase III have been located by hybridization and sequence analysis in the small unique region (Us). The restriction enzyme mapping and hybridization studies have revealed the existence of deletions in different strains of EBV (Pritchett et al., 1975; Sugden et al., 1976; Hayward & Kieff, 1977). The B95-8 cell line was established by infecting marmoset lymphocytes with EBV derived from peripheral blood leukocytes of a patient with infectious mononucleosis (Miller et al., 1972). B95-8 DNA has been shown to be missing approximately 13.6 kb from near the right-hand end of the long unique region (Delius & Bornkamm, 1978; Raab-Traub et al., 1978,1980; Bornkamm et al., 1980; Heller et al., 1981). Part of this deleted region has been shown to be homologous to a region just to the right of the large internal repeats (Raab-Traub et al., 1980). The significance of the deleted DNA is unknown since, so far as we know, the B95-8 strain shows all the functional properties of other strains not carrying the deletion. However, two components of the EBV membrane antigen complex (gp350 and gp220), which are antigenically related, are expressed in a uniquely abnormal ratio in the B95-8 line (Thorley-Lawson & Geilinger, 1980).

We are in the process of determining the complete DNA sequence of EBV; here we describe the DNA sequence analysis of the *EcoRIC* fragment of B95-8 DNA (see restriction map in Fig. 1) and the identification of the deletion point by comparison with sequences determined in Raji EBV DNA that does not carry the deletion. We present a detailed analysis of the possible coding regions and potential transcription sequences found in this region of EBV.

Materials and methods

(1) DNA recombinants

Recombinant cosmids containing the *EcoRI* C fragment of B95-8 and Raji EBV DNA were obtained from B. E. Griffin and J. Arrand, ICRF Laboratories, Lincoln's Inn Fields, London. DNA purifications were carried out as previously described (Birnboim & Doly, 1979). Recombinant DNA inserts were isolated by restriction enzyme digestion followed by electrophoresis through LGT agarose (Wieslander, 1979). A random subclone library of the *EcoRI* C fragment was generated by sonication (Deininger, 1983) followed by cloning into the bacteriophage M13 mp7 (Messing *et al.*, 1981), M13 mp8 or mp9 (Messing & Vieira, 1982) vector systems.

(2) DNA sequence analysis

Single-stranded DNA templates prepared from the M13 subclone library (Sanger et al., 1980) were sequenced by the dideoxynucleotide chain termination method (Sanger et al., 1977; Sanger & Coulson, 1978) using a complementary synthetic oligonucleotide primer (Duckworth et al., 1981). Sequence data from the M13 subclone library were aligned and overlapped by computer (Staden, 1982). The complete sequence of both strands of DNA was established. Completion of the sequence required some non-random sequencing. Initially, M13 clones were not obtained surrounding an Eco K site, which happens to occur in the EcoRI C sequence ($Escherichia\ coli\ JM101\ is\ r_n^*$). Clones covering this region were generated by isolation of restriction fragments covering the site and forced cloning into M13. A total of 536 sequencing gel readings were performed to obtain the EcoRI C sequence. The total length of sequence read was 112,581 bases. On average each nucleotide of the sequence was determined 6.5 times.

(3) Nomenclature

Because the large open reading frames and potential transcription signals are found scattered on both strands of the DNA we have used the nomenclature R (rightward) strand and L (leftward) strand. Reading frames on the R strand are named RFI-5 and those on the L strand are named LFI-6. In order to avoid confusion, in the future we propose to prefix these names with the initial letter of the restriction enzyme and the fragment number, e.g. EC – for EcoRI fragment C; BX – for BamHI fragment X. The in vitro promoters are designated R or L (for rightward and leftward) followed by a number. In order to avoid the confusion of using two nucleotide numbering systems for the R and L strands, we have used the R strand numbering for both strands throughout. Although most groups use a map of the type shown in Figure 1, we note

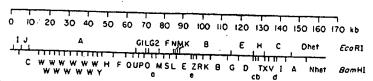


Figure 1. Restriction map of EcoRI and BamHI sites in B95-8 EBV (Arrand et al. 1981). Scale is in kb.

that a few authors prefer an opposite sense map of the EBV genome (Hayward et al., 1980).

Results and discussion

(1) Sequence analysis

Because of the lack of genetic data for EBV, we have adopted the approach of first establishing its DNA sequence then using this in a predictive fashion to design further experiments to understand the coding and expression of the EBV genome.

As a first stage in the detailed analysis we have determined the DNA sequence of the EcoRI C fragment (Fig. 2). Marked on the sequence are major open reading frames, potential promoters and splice junctions, potential poly(A) addition sites (AATAAA) and the three promoters which were detected by in vitro transcription (Farrell et al., 1983). The organization of the reading frames, promoters and poly(A) addition sites is shown in a simplified form in Figure 3. Although smaller coding regions, whether genes or individual exons, are possible we have excluded those open reading frames lacking a possible initiation codon and under 200 codons in length. We have also examined the codon usage of the reading frames to see whether they have a distinctive preference in order to try to predict coding and non-coding regions.

It is difficult to predict the function of a DNA sequence without any other biological information about it. This is especially true of eukaryotic sequences where complex RNA splicing may occur (for example, adenovirus, reviewed by Tooze, 1980). Although there are specific sequences associated with transcription and RNA splicing, these are consensus sequences and any particular signal sequence can be quite different from the canonical sequences such as the TATA box (Corden et al., 1980) and the CCAAT box (Efstratiadis et al., 1980) of eukaryotic promoters. In some cases nothing resembling these sequences can be found in eukaryotic promoters, e.g. SV40 and polyoma late promoters and the adenovirus E11A and IVA2 promoters (see review by Shenk, 1981). There are consensus sequences for splice donor and acceptor junctions of AG/GTPuAGT and Py, N Py AG/G, respectively (see collection of sequences by Mount, 1982). The redundancy found in these sequences makes their prediction very difficult. Nevertheless, when these promoter and splice sequences are near to their consensus sequences and are present in logical combinations with open reading frames and the polyadenylation sequence AATAAA, then their presence can be highly suggestive of transcriptional units and protein coding regions.

C166C1C8	LICCTICENTIC TICVESENCYCCUERSTISICENCY	CLINCTICCNICCY	. FECEGEA IGENE	LECTALIER !	10
**************************************	TICCAIGEAICCTICAGAGACAACAGGBATAICGACA 50 40 10 40 ARCGTACCTAGGAGGCCCCTAIACTGC 4 5 5 5 7 4 7 1 0 7	EARTCTACETACET	AGGCCETACCCC	CCATTACIEA	CTTANCTICEGO
TIGECTET 180 AACGGAGA A E R EC-LF:	ECEGTTAASCATACACSCCAAAICTTCGGIGACATT 140 150 150 150 150 150 150 150 150 150 15	13113A411AA61161 130 130 	ACCABACETOTIO	1466 44668 14 1 10 1 16 1 16 16 16 16 16 16 16 16 16 16 16	CGCTCCCETTA
	TOTOCATEGEATECHTOTALTECHAREATICCCCCTCT 200 240 250 250 250 250 250 250 250 250 250 25				
TABICCCG 248 AFCA66GC - 9 8 F (m4 EC-LFS	RACACCREECRAARANCACACACACACCRACCCRACICCTA 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 130 13	318 318 4 4 6 8 6	CICCEACCETCEC SOO CICCEACCETCEC	44(CTCGCTG 005 116636764C	ICCYCECLES ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCETERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCECTERY ICCYCET
SCCTCCAT 488	CARAMAC IERCA ICC I AT IECCECCATICI ICCALTEC. 130 410 410 410 410 410 410 410 410 410 41	**************************************	A6A 1CAC(116(1 110	CABARTCAGA MO GETTIACION	ATTICCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
CC-LF3	F Q S A I A P A E L A	L 4 A 1 I	µ	• •	
	ATCCACAMACACAACISAAGGIAGGCTAGCGTICGG OO 510 520 530 CAAGTGTICGTGTTIGACTICCATCCGATCGLAAGCG D T (T Q L T A L T Q				
ACAGCET 630 161C6A4 L E [C-LF6	GACTICC TIMETACAC FAMEST GCTG TTCTTATAC 50 ESG GEOGRAPHICATE TEATTICTACCACAAAAATATAC 7 C L L 7 S F J B C T	1	10 10 10 10 10 10 10 10 10 10 10 10 10 1	61466116118 61466116118	COCASEACAAAC SSO ECRTICIBITIO A L C F
. IS becomes	IE HERIEALIELIETTAGACCGGCACATACCTICACA	SETTIGATEACATA	1416666616166		
	TERECECTIAL PROPERTY OF THE PR				
610 610	TABEL TO TEMPERATE CONTRACTOR OF THE CASCASCA TO CONTRACTOR OF THE CASCASCASCA TO CONTRACTOR OF THE CASCASCA TO CONTRACTOR OF THE CASCASCASCA TO CONTRACTOR OF THE CASCASCA TO CONTRACTOR OF THE CASCASCASCASCA TO CONTRACTOR OF THE CASCASCASCA TO CONTRACTOR OF THE CASCASCA TO CONTRACTOR OF THE CASCASCA TO CONTRACTOR OF THE CASCASCASCA TO CONTRACTOR OF THE CASCASCASCA TO CONTRACTOR OF THE CASCASCASCASCASCA TO CONTRACTOR OF THE CASCASCASCASCASCASCASCASCASCA	1 2020 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2	TAGGA4CCAGGCA 150 ATCC1CG1CG1	14 14C 161G 1T. 140 141G 164C AA	erecicce i i i Ciconericava
£ £0.415	* A B * Y S E L Y 6 A 9			. 1 1	CAERAIRITI
	THE THE TENEDRAL STATES OF THE				
COACAE 990 ISCITE 5 L EC-LFS	CTCCCCACCLECATIACACCCACACACCACCCCCCCCCCCCCCCCCCCCC	TAGGAGAGERGAGE TO \$40 A FCE FETC HERAC F S L 2 Q	LA FARAGICACIA 138 TTATTICAGIGALI	CAMARCTICAL BERTSTEELS CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCTICAL CAMARCT	CATITECACCES
AGAGAC 1046 ICICCG	CCACCAGACICEETRACCAGCAAGETICCACIICCTAG 0 1950 1060 1070 CGTGGGCGGAGCGAGCAGA A G S T V L L S B E G	CCGCGCTGVC19EV	CAACAATCC 656) 1070 14 11 61 1460 1661	16(AAAA700 1810 14(61111AA4	431462641CA 0001 134736334111
i à cc-Les			WAILICATCICC		*******
CACATF 1370 81614A Y B [C-LFS	GCCGACGESTCETECCETECCETGCCEAGGEC O 1148 III-90 III-90 COBCTOCACARAAGOOGAAGGETCCAGG A S F Q E W W A D F F W	ITACEAFTTTTEAC		CCCCTCCTT	TATE TATE
	ATCCCTTCCCCCALCTACCATCCCCAACAA 1200 IZAA 1200 IZAA 1200 IZAA 1200 IZAA 1200 IZAA				
N E [C-[/5 1360 Tac Tab	ANANT MEDIANCACE PAGE 17 IGN FACTEGIACAT 910 1330 1330 1340 FILTRICCCI I GICCAIC CAMACTAT GACERIGA F L S C B P R F R R	1300 13	L E 8 8 8 CVEVVVERSIRES	TCGGATCATC	1270 1461114C 14461
EA446C	TORCE I BE O B O C HE C I BECC I GECCOCACE (COLORADA TORON COLORADA TORON CATEGORIA DE	CETTCC1C1846CC	4CCC&CA7&CICT	4C#4118GA40	1146416641CA
FCGAA 1930 AGC 	ACTIALIAAACIGGGECCAIBBECEGCCABBIBI FO 1500 1518 IGAAIBAIITBAECGGGBGIACCCGGGGGCCACAA E I E E B A L L H E	CABCATECTIEGA 1180 169 07CGIACEALCCIA A # E \$	1430 1430 1430 1430 1430 1430 1430 1430	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	A 4 5 C
AACTC 1474	EREAMERGECCIECEGAAATTABBCICATCCICAA 1548 1400 TESE CICIICACCGGACCECCIIIAATCIGAGIAFGRETCI 5 3 7 8 8 7 8 5 5 5	ETITIAACECE EAA	CACCATICITIES	11466611466 1150	64CC16AA6GC11
Ciéis S C (C-LFS	CICITE RECEGGACE ECTITA A TERRETE ESTE	-			6176
ECACC 1/10 C6166 A 6 (C-L/5	TTTCTGGTACATGAAGCTCCCGCGTGGTGGAGTCCGC 1640 1400 1700 1 SANGACCATGTACTTCGAAGCCGAGCACCTCAAGCA R Q 7 R F S B- S T S Q R	G C C C C C C C C C C C C C C C C C C C	611616050616 611616050616 611616050616	CATCGAGCCC 1640 6140136666 0 L 6	1636 1636 1636 16461 1636 16461 1646 1646
1100 (AICC	GCTGGGGCCACACTCCCGGCTACGGGGCTTCC8866	64867C164C81C 6736 1766 6CCCA6ACCCCA67	CCLLECICCICLI Tive Tive	AASI 11 18CG 1130 TIÇANAACSCI	171016CACEA 1720 1730 1730 1730167A
	centricecelecticus coecelificateic	egarecercic;	ETECTICAÇUEACTE 1 8 29 CAÇCAÇUEACTE CAÇCAÇUEACTE	eciricocci	**************************************

Fig. 2

1900 1910 1910 1920 1930 1940 1933 1940 1930 1930 1930 1930 1930 1930 1930 193	
	((.1)
COCCACCIDETABABECCCIC PICATCHICCICACCIGCCCIB DE ICACIACCABITEGEGIAABAC IBAGEGGGAABACICAICABA 1990 2000 2010 2020 2010 2020 2030 2030 203	t c -u
ETCACCETALALETTICETGETTICATTICAGGGGACTERICCTCCTGMATTITCGCCAGCCGCCCGGGGCCGGCCTCCTTTCC 2000 2100 2100 2100 2100 2100 2100 21	
> 5tert et-et1	(5-(6)
STANT EZ-BEI TEGNANICENCEATEGNICEACETEGGICTOTTTOCCCTTTCANATHCARCTTCANANTTTCACASICTICCECCACTARAN 21 18 21 21 21 21 21 21 21 21 21 21 21 21 21	CC-RF
Start EC-175 C)[
CCCCCTCCATTCGCTCGCGTGSAABCCACAATGAGGTTCGACGAGAGAATTCTCGACCCTCGTTGACCGGAGAAAAACAGAGACGG	(C-AF
CTCGGG ACCTACGO ITACAGACTT AGACTATT I I CGAGATTACCT GGGT GGGG A ICQUANTIC CONTROL TO CACCT	(C-0/)
TILICCCEGE TELETIFICATE TELETIFICACE CONTRACTOR CONTRAC	fc-efi
CONTRACTOR OF THE PROPERTY OF	(C-61)
CHIEGE ALTHAMACACCACCACCIACACACCACCACCACCACCACCACCACC	EC-4/1
Start (C-872) L H L S 4 4 8 6 A L 8 A 8 L B R P 8 P 1 A 4 E H 8 P 8 CCIGCIANACCICCITECTOSCAGGAGGAGGICICITICATRACCICCCCCCAGGAGGACCAGGACCAGGAGGACCAGGACCAGGACCAGGAGACCAGGACCAGGACCAGGACCAGGAGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGACCAGACCAGACCAGACCAGACCAGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACCAGACAACA	(C-815
THE TERM OF THE THE THE TERM OF THE TERM O	(C-8f2
C A F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F G P C F	EC-RIS
CICCTITAGCOSTCCIGGACCIACACCACACCACACACCTTCGACCACTCACTCACACCACCACCACCACCACCACCACCACC	10-111
CHIEGEC.ACCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCC	[C-8F2
BORGECCECT TEATISET RECECCECTAGECTECT CANCER CT TESTALATE LET ACCORATE TACCES CANCER CECTACE CT TESTALATE LET ACCOCATA THE ON THE ON THE OTHER CHARGES CANCER CONTROL THE OTHER CHARGES CANCER	£C-4/2
ACCCIGETACAGECCATECOGTANISCATIGCAGCOGCCCCTRITCCCINACCAGCTGAGCGGGCAGCGGAGCTTACCAACTCGG ACCCIGETACAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	(C-AF2
TO THE TOTAL TO THE TOTAL TO THE TOTAL THE TOT	(C-4F3
ectaccitanterieres ractificiacitecitecitecitecitecitecitecitecitecite	EC-1F2
CICEACT LOCCCECEALACCTAGEAGAACCATCECTACC AGECCAACTCACCCACCCAACTTTTTAGETTTTAACCACCC 3510 3510 3500 3600 3510 3510 3500 3600 3600 3600 3600 3600 3600 3600 3600	(C.arz
R T E R Q 3 L A P L R Q A A T P P C 1 P B R L T R L MACANGEAMAG GGACCHATCCAATAGCCGTCTCAACCGACAAGGAGGGGCATATCCACCGGACAATTAGCATTTACAACC 3610 3670 3870 3880 3840 3840 3840 3870 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 3880 3870 387	(C-B15

Fig. 2, continued

TETTISTESSACE CECATIACTES CECACACACCECCECCECCACACCACCACCACCACCACCA	E [[.4#
CCLACCIALITETISECCESCECATE GENERALITETISTE CALCALITECA ACCIDATE TA CALCALITECA ACCIDATION ACCIDATIO	:
TOTT THE STATE OF	4 (C-4F)
GACCAMECCIACIA CACACACACACACACACACACACACACACACA	W (C-11)
GESCENTALACETRACECACCETACECACCACCACCACCACACCACCACACACA	L SC-RF2
ICAMETICATECCETICAGEGACCIATICECATAMECIATICACCEACCMACTICACECACCACCACCACCACCACCACCACCACCACCACCAC	
ACCITITATION CONTROL TECCHICCATECTECCACCACCACCACCACCACCACCACCACCACCACCACCA	t (C-m/1
ACCCCTECCCA ECT TO COCCCCC COCCCT COCACC TO CATO TO CATO COCCCT ACCT COCCCT ACCT COCCCT COCCCT COCACC TO CACC	(C-#1)
AND EC-MET) CONTROL	
4444) TTILACCICLET CONTICUENT CON	
CCCTGTIMITCETICGECATICGMAS GRATGGCACGETCCTOMMELGRATAGCCGGCCCAGCCGGGGGGCGCCACAMREAGGGCCCAAGG 488 4418 4418 4418 4418 4418 4418 4418	
COTOCCOLARCEONETTICATECTICATECTOCACCASTOGACAMECANETAN AND AND AND ATTO ATTO ATTO ATTO ATTO ATTO ATTO ATT	
CLAMBATTRAMENGCAEATTTCCETCACCACCETRACCTOCTCLECTCLECTCCECCACTCCARGICCLARCTCCACACACACACACACACACACACACACACACACAC	
ASAGEMENTE PSACT CASSECCE CATECAMANT REAL ANALASANCE TO RECECT TACCAS CATEGORIS CONTROL TRECES CATEGORIS CONTROL TO ANALAS CATEGORIS CONTROL TO ANALAS CATEGORIS CATEG	
CCACCICACCACACACCATAMACAGCAGCACCATATICATCCACCACCATCACCATACACACACCATICATACCAGATICATACACATACTACACACATACATACACATACATACA	
TEACECTTCCCCCCCCCATALACTIALACTIALGETIALGETIALGETIALGETIALGETIALGETIALGETICALALACTICALCTCCCCCCCCCCCCCCCCCCCCCCCC	
GAGCACHAACCGGCCAGTICTTTAGCATCTICAGGAGGAGCCCCCGGAATCCCAGCTGTCATTCCCGCGTCATCCCCGCCACCCCAGTC 140	
SE OFF EC. OFF SEC. O	£C-8F3
ICETTE CONTROL DE LA CONTROL D	[(-BF3
codericare i rice de la contractica del la contractica de la contractica del la contractica de la cont	(C-4/3
APPIRACEATRICCATARICCATARICACETACCCICCTCACAACCCAACCCAACCCACCCTCACCACCCAC	(C.8F)
uterfact act rice treate (na actiate canace constant and a same to act test in actiate of a traite for a Great the constant canace contillar account for a continuous action action action action and a continuous and Great the constant action account for a continuous action ac	((.4/)
CISCOLICA CONTRACTOR C	(c-er)

Fig. 2, continued

A Q I C R D A R A A C R C C C C D N P L P L C C C CECCEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAITEAGAIT	((-#)
ti francter	
ACCANICATICA TECNT TOTE ACANOCICE CICANACI PICTACE CON CARRA CARRA TECNTA COCANCE ANTECIA TOTAL A 1840 SERVICIA COCANCE ANTECIA COCANCE ANTECNICA COCANCE	· (C·H3
A S D A P D C Q E P D E A C Q S P P P A S I D P A T S I S ENTRE (C-L A ESCLARECIA CECCCECIA CALAMASICE GALAS ESCLATIFICA DE COCACCIA CECCIA CECCACCA CACACA ENTOCIA TICAL 3 19 10 10 10 10 10 10 10 10 10 10 10 10 10	IF1e EC-RF1,1e
HACCAA TACCACAACAACAACAACAACAACAACAACAACAACAACA	£C-RF3,34
TOCHTCACCCOMCTOCAMCAMCAMCONCOMECCACHTCCCTTACCCTTACCCCCCCCCCCTTOTCCCACCCCTTTTCCCACCCCCCTTTTTCCACCACCCCCTTTTTCCACCA	((-4F),14
TAPS TOPE ACAL VAPACE 11 TSPECT FAEPACE CONCECCECCECCACCECCULA CONTROL	(C4f1,3)
COCCOCCCTORECCCIACONACACCCCTCCCCCACCCACCCCCCCCCCCCCCCCCC	EC-4F3.3a
CECHATIFECCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	EC-4F3,3a
CENTRE RECOCCECTE ACTIC HE COCCECTE ALACCECT CE CECACACE CE CECACACE CON CACACE CON CACA	{C-4f3,3a
C S T H {	EC-RF1 34
ACCAGECTECTECTECTECTECTECTECTECTECTECTECTECTE	CC-RF3,34
CACCACACTIC TEACCACCECACCACCACCACCACCACCACCACCACCACCA	EC-N73,30
CALA A B 1 C P P P P T P P P P P P P P P P P P P P	CC-RF3.3e
ALTESTERICE CITECTO CONTRACT PAR 3 P. A. C.	((-H/) 1s
ICCCLL R L & V A .	
C E L N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E N E	EC-174
CHIGGECCECCCCCCABCCCCCCCABCCACCCCCCABCAGGAGATGATGATGAGGAGCACCCCCGAGGCCACCAGAGGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	ec.ura
SECREGECCECTECACCEAGETGETCTTAGGGCCCCAGMATEGATTGEMAGGCCCCAGEGTTACTGCTTATEGTGGCCCACTTT 7718 7719 7719 1710 1710 1710 1710 1710 1710	(C-LF4
ALEMACAGECCECTATGGGGECCTSTGACTGGTGCTTGTGGTGGGGGACTALTGTGGTGGGGGCCTATGGTGCTGGGATAACAGT 1300 1310 1310 1310 1310 1310 1310 1310	15-114
AMERICA I ACECTEAC TERRESE (CERTICES CATESTA PICE (CETTETT COCCE OF TRANSPORTANT FRANCESCA CECCER 23 Me 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240 / 240	EC-1/14
ACTECAND TOTECHOTC TEACH TACAGE TO THE THACE TO CATGE CANADCATE TO THE GOAL THAT CASE CACAGE DOTTO THE AT A 150 1700 1740 1750 1750 1750 1750 1750 1750 1750 175	EC-114
SICATIFACCIA TETROCETAGACIA TALTETAACCCCCETTACTTA TATOACTICCETTOTTEGAGETACTICCATOCACACIACACIACACIACACIACACIACAC	EC+1/4

Fig. 2, continued

TTGTCCATCCCCGGGAGTGTATCCCCCGCCCAACTCGATCCGACCGA	
THE COLLECTION OF THE PROPERTY	€C-L14
ACCTGCATGGCGGTGACATTIICAAATITAACCIGITCIGAGAAAATEGACAAACIGACCCCACACACACACACACACACACACACAC	
ACCIGEATEGECEGETGACATTITCAAATTAAACTITAACCHSITCIGAGAAGATEGACACAGCACCACCACACACACACACACACACACACACA	CC-184
A TEASTEL THE CHARGE OF THE ANALYSE COURT OF THE PROPERTY OF T	
A TEMETICA FILE TAMACI DE CAME I EL TAMACI DE CECCICO I EL CANGECET A MATATE DE COMPOSE A TEMETICA CI I I CATALLE CONTROL DE CONTROL	
(Stort (C-LF4 H	
I BECANEANCAMECENTACCES BANGA BENGE I BECCHETI BECCAT TITTECANGE TGE I NANCENES ANTRE PROPERTY AND ANTRE PRO	
ISCARCARCACCATACCCERAGAGEAGCISCCOGTISCCATTITICAACCISCIAANCACGAGTAGCAGCAGCCAACACICCT 130 1950 1950 1950 1950 1950 1950 600 600 ACCITETIGICOGTATOGGCCTICICCTCCAGGGCCAGGGTAGAAAGTTCGAGGATTIGGTCTCCGICCGGTTCTCCGGATTCTCCAGGA	
CARCARCA IBRAGACCICCALAGERANCI BECAGAICLAGGARICACGI AGECACIAGGGICI CA IGAGGIAACAGAICCCIAAAA	
CARCARCAIGNAEACCICCALAGANACIAGCALGENECA ACGARICACCIA GECCA INSCELE ITECATA CARGARICACA ACAGA TECCA ACGARICA CARGA	
ATBECATCACESTATE TAAACCEART TITECEGGGCALCESTCACCCACACACEGGGGGGGGGGGGGGGGGGGGGGGGGGG	
ATBOGATCAGGGTATE TRAACCOGGT I TICCGGGGGATGGGGGAGGGAGGGGGGGGGGGGGGGGGGG	
BEGGGAAAAAAACACCAGCTATAAGGAATTGCIGAGGCCAAAGTIBTTCCTCAGGGGCTTAAGGCAATTGGGGTACCTA	
GEGECLALBIRTACACCACCTATAGGRAFTEC/CLAGACCAAARTIBITECTC LGGTGGETTIABACCTAATRIAGGCAFTGGGTGCCT 100 07:0 07:0 07:0 07:0 07:0 07:0 07:0 0	
GAACATICCIANIGIGCCCIGGGTTICCCTCCCTTCAICTANATATICIACCTCCCCCGCCTGGTGCAANATGTCTGCCTCACAATACTA	
CACCATECTAREGICCC INCATTICC FOCCTICANCCANATATICIAC CICCCCC COCCIGETECAMASTECTACCC TCACCATECT 1740 8380 8310 8310 8310 8310 8340 8340 8350 1370 CTESTANCESTICACCOMACCCAMAGEACOMAGIACETT TATAMENTERICORECCCCACACTITICACAC OMAGICITATORA	
ACACCIANTCCANCCTANCATIC TATCACI MACCEGGAGANACTERTANGGACCEGGGACTI TEGECCTCCGCGG TETETCETERET	
ACACCIANTICANGCIANCATICIATE TATCACIANACESCARANACTRITARAGACCECEGAMITITECECCTICAGGG TOTICAGISETCE \$100 8190 4400 1410 8070 8430 8430 8440 8460 4600 TETCERITARCITECANTETANGATAS TENITITACECS (CTITIRRETRITIC TOGGGCC TCANACCEGAMACOCCACAGGCCACAGG	
TEACHER TECETICE CECECE CECECE A LOCAL CONTRACTOR CONTRACTOR TO THE TANCE OF THE PARTY OF THE PA	
TEACAMETECCETCCCCCCCCCCCCCCCCCCCCCCCCCCC	
14:14:15:14:14:14:14:14:14:14:14:14:14:14:14:14:	
THE THE PART OF TH	
CIGINAL ACACCACCIA MALCA GOCA I SCII MACCII I SEFENI CITARICI TARICI TAR	
CTREINACEMECCACCIAMENACAMOCCATERINACCI FICETEMACCIAR INTITARINITARICIALAR FARRICICEMECCA 8-84 8-84 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-870 8-8	
[[]]	
4878	
£41774474444444444444444444444444444444	
THE STATE OF THE SECOND PROPERTY OF THE SECON	
BECBICICETCACEBOABAABAGTTAAC AARCCEICEGECETTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	
BEEDICTECTECECOMMENABRETTANCAMECCRICEREEDICT FRETEGAMETRECHECTSTEAMIACROCECCIACECECIA 8770 8970 8970 8970 8970 8970 8970 CCF MARGANETERECCTCTTCCCCAMTETTIERRCALCCCCCCAMATERECTTCACTREACRACACTERRESCCARRESINERROSCCUT	
TECACIAIGICTCACCCACACTATATOTCACCTTACCTCCCTATATAAA	
TECRCIALECTOCCCCCCCATATOTCCCCCTTACTICCCCTIALIANT CACACCCCCATCAACCACCATCAACCACCTATCAACAAAAAA	
INTEGRACALCHIC PANTETE DECERCAL BENEGENIACH TENCERCE TRANSPORT THE TANCETON TO	
FAICCARCAICINC INACICTERECCEC II SCACECAIACA IEACCAC CIRANGIC TRITIANCERER ICAC TERTITI TE FIICCOTERATE 1100 010 010 110 110 110 110 110 110 11	
AGACACAAGAC IGCCAGCCCCAT FEEGGAAGGGGGGTGGGTACGGGGGGGGGTTTTCGCCTCCTTTAAACA	
AGRACIAMENT INCCASCOCATIBACA MAGGORATEGET AL MAGALET TI EGECI (ETITI MAGALETCI CICA I CIGATECI CICAT 9180 9200 9210 9210 9210 9210 9210 9210 921	
#95-6 BELETION POINT	
1180 1140 1100 118	
STATE TO THE STATE OF THE STATE	
314 3180 4144 1100 414	
11 1616ETACTETCCETCCGGCAGGETBAARTGGBACACAGTTTARABAACATTARCCCTCGGGTCCCTATGTCCACCCCCAGGCCTGTETCT 1918 1920 1959 1000 118 1410 1410 1450 1460 1460 1460 1460 1460 1460 1460 146	
9468 9478 9480	
STEANACSTEECGETAGGAECGTGAGTCCACCTGACCGATCCGTGGCAAGACTTCATAGACCGCCACTGACAGTGACAGTGACAACTACAACTACAACTACAACTGACAGTGACAACTACAACTACAACTACAACTACAACTACAACTAACAAC	
CAB TITECHELECCATECTICACH TRABCISCALTECTIAGEC NCC TITTEANBTATETICICE TO TETTICACTICA TOTTEANGRA 8 TOTAL 10 TOT	((.11)
#ICC+1AAA+168C+CAAEC+CCAECCETA+ACTA+CA+14+14+14+14+14+14+14+14+14+14+14+14+14+	
CACCATIANAMISSECTEANECTECANGCETETATAFTATATEACCACECACCANACANACCACCACCACCACCACCACCACCACCACC	
A A A B B B B B A A C B B C C C C C C C	CC-LF\$
ASACTCISTSCHATCANACCCCASSCISCCSGCCTCACCCASSASSA	
TOOM 9650 9660 1110 1610 9650 1700 9710 1700 9710 1770	
AGRETCIETACEATCASACCCCAGECTECCAGECTICACCAGEANGACTACCGCCAGEANGACATABACCAGAGGAAGACCCCAANTGAG 110 - 1400 110 - 1400 111 - 1400	(6-11)
TCCCGTCAAGCTGACGTTGCAGCTCGGCGGCTTCCCTCCC	
ICCGGICAAGAIGAGGITACAISICAGGCDCGGCIICCIICCCAAITIIGIGICGCAAAAGITIICCAGATEAJGATAACTGICAGGCICAG 7738 9738 9738 9730 9730 9730 9730 9730 9730 9730 9730	
T T T T T T T T T T T T T T T T T T T	

Fig. 2, continued

A. T. BANKIER ET AL.

CATCAGEACIGC CGCCAGTAGGTAGCAGCCGGGTTTCCAGTCCTTGGGCGCGTGTGTALLCTTTTCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT	
CATELOGACIOCCOCCAMIAGRACACCOCCITITAS ICCITOGAC GEOCACCATE IGEACACCE I IGE TEGEOCOCCITICA (ICCITORA ICC 9820 9820 9820 9820 9820 9820 9820 9820	
660GCCTCTCTCTGABAT TOGCCTCGBGGCCCATGCTANTGAGGAGGATGATAA TCAGCACGACGACGACGACGACGACGACGACGACGACGACG	
COCCCCAGACAACTITACCCCCCCCCCCCCCCCCCCCCCCCC	10 17 (C-L)
CAGGCACACATTTCTCTGGGGGGGTTCTCTTTGGGCTGGACALTGCTAGGGTCTTCTTGGGGTGAGGGTT	
CAGECACACATTECICIOSEAGESTICICITISECCESTICOCETICAACASTECIAECETICITCICCAACECCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACCATACATACCATACCATACCATACCATACCATACCATACATACCATACATACCATACATACCATACATACCATACATACCATACATACCATACATACCATACATACCATACATACCATACATACCATACATACCATACATACCATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACATACAT	κ 10 16 1 ((-1)
GEAGCACATCARGCCCAACAACATCATGTTCISESTI BALANCTACACACTCC TACATCACTAC	
GENOCACATENDOCCEMENTATE TO TOTAL SEGMENT OF THE SEG	
CARCAGEAGITE FATGAGMEAGCIATEGGECAGICATECATTEGGTGAACCAATTITETTICAAAGGTGATAACCAATTITETTICAAAGGTGATAACCAATTAACCAATTITETTICAAAGGTGATAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACCAATTAACAATTAACAATTAACAATTAACCAATTAACCAATTAACCAATTAACCAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAACAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAATTAAAA	
BARCABERGITE TATBAGUSEACCIATE BECAGGEATECA FIESE FRANCA CAPITAL CARGA FARAINA CARGA ECETE FEET. BIND 10100 10200 10210 10710 10710 10010 10100 10200 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210 10210	C
1617ACCACAATCACCACCACCACCACCACCACCACCACCACCA	
TRETTACCACACACACCACCACCACCACCACCACCACCACCAC	10-11
GACOBANGICA IE I IGOCCACCAEGO I CCCCACGGIGGACGCEGGGGGCCCAIGG I GGARAGGAIG I IGCI EGICAGACCI AFAI I GGGI LA	,
CACRBARATER TO TECCTACCARGE TECCTACCATCGATCGATCGATCGATCGATCGATCGATCGAT	[[-1]
IBICIACAGEAAAAGAGGAACTIGCCCAGGCAGIGAGTIATTIIGCAIGCGACCIGCCGGCGGGGACTICCAGACGTTICTCCAT	
TRICTECACIANACAGEALCTICCTEAGGCAETEMETATTITICATICTECACTECTECCTECCTECCTECACTACTACTATTICTECAT ACAGCCATECTTECCTTACACTACTACTACTACTACTACTACTACTACTACTA	
TECCECTTACETE TECATET TECHNEMERACIAE CATTICE ACAMANTIAN CERRACITECATA AT TECATICA TECATICA CONTROL CENTRAL CONTROL CON	
PERMISSIFICATION CONTRACTOR TO A TON CALCACTER CONTRACTOR CONTRACT	
INCATESTICCAC COCCUTICCCONTESTESTICS COACCORDETESTISSACCOCTOSTISATATICESTIGAGATICTESTIGAGASCET 19450 19460 1959 19689 19689 19650 19680 19690 19700 19710 1945TACTAGGETGREGRAMGORETACTACTACTGTGGGCCGACCACCTTCGCCGACTGGTTTAGCCAACTCTRAGCTACTCTCCGA	
ROSSEGGICTITGCCCTCLIGGCCRAGGCTCCTIC ICC LAGGARTOCCCABCCCCCTBCACTACCTTCGCTCCACTGGGGATCTTTBCCA	
RESESSO (CT) TOCCCTCLASCCARGECTCCTTC ICC LABBARTECCCARCCCCCTRCACTCCTTCCTCCCATGESSCRICTTRECA 1978 1979 1974 1975 1975 1978 1978 1978 1978 1978 1978 1978 CCCCCCCCARAMACGGRAFICCGCTCCARGAMGRAFCATCCTTACGGCTCGGGGGGGGTGTTCGAAGCGRAFTGACCCTTRAAAACGGT	
CC ITERITRET BRATTER BECRATICE CECER TRAGBATIC TEGECRACE CECER TECT BET BLECC BET GENERAL CECERCE CECER TECT B 1 00 10 10 10 10 10 10 10 10 10 10 10 10	
AGGECCECTTECTTICCCTREACTTIGGGGATATICEGABTCTGCTCTCCCCCCTCTCTTCTCTCTCTTTAAACGATATATTATTT	
AGRECCCCTTCCTTTCCCTTGACYTTGGGATATTCCGABTCTGCTCTCGCCCCTCTCTCTCTCTCTTTTAACCAGAAATATGTAGTA -10180 1016 1018 1410 1616 16150 1816 1816 1817 1618 TCCCCCGGGGATAGGATACGCTGAAACCCCTATAACCCTCAGACGGCAGATAAAACAGCAAATTTGCTTCTTTATCATGAT **********************************	
1099 1006 11010 11020 11030 11040 11040	
GOTCHAIRTEIGABECCCCTCAETTTEGETCTTAGAATGGTCGCCCGGGTTGTAMATTCTAAAGGCCCCCGAATGGGCCCCCCCAGTTG 1000 1100 1100 1100 1100 1100 1100 110	EC-172
11890 11899 11100 1110 11100 1110 1110	
APTENDAMENTAL CONTROL OF THE PROPERTY OF THE P	£C-1/2
THE THE THE TEST OF THE TEST O	
#THERECEASTICETICASCERTETECTICITETECCECCTICATECCCCCTICASCECTCCAMARRESTATICATECTOCOSTICATEC 11:00	£C-(F\$
CTBREACAST FECTICETICESCICOACEAACTICISCIA FACCESCIA CECENICISCES CECTICISCISCISCISCISCISCAS CONTRA LIBRO 11310 11310 11310 11310 11310	
OPLE (RETYFERENCECTICIACCOCTCCACCCCATAGACACACCCCACCCCCCCCACCCCCCCC	EC-112
eacticoste arracteracicocccateggaco escutacos de la 1900 de 1900 esta de 1900	
THE REPORT OF LANGE TO STATE	EC-113
ARECCCACAGGTAGCCCTTGTBERARAFGACTCTTGCARBCCCCGGTGCCAGRAGAGCACCCCCCCCCCTCCTTTACCCGGGCATCCGCC	
F E F P L G T F S [B F C C C C C C C C C C C C C C C C C C	((-1/2
ACCARD FORM CACCECCECCOCCOCCOCCOCCOCCACCACACACTIGGAGGCCGCCCTTCATCAGAGCTCCACCCCCTTCATCAGAGCTCA	
ACCABBICABACCACCICCECCECLACCACCACTACTACACACACACACACACACACCACCACCA	16-172
TECGTCAGEACCCCCACATATETCTTTGTTAGCAGCAGCAGCAGCAGCAGCAGCTCTCCCGGCCTCCAGAGAAATCCCCCCCC	
PECSTEAREACCCCCCCCACNTAPETCTTETTTGTAATCSGEATCAGCTACCAGGAGGACCCTTCCCCGGGCCTCCAAGGAAATGGGGGGCCACA 11850 11850 11850 11860 11860 11860 11860 11860 11860 11860 11870 GGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	10-113
MANUFACTOR TENTO TO THE TOTAL TO THE T	
ABBOTTCCGGOTEGTGTGGCCCCCCAGCCCATCGGCAAAACCCABGTCTCGCTCTCTGAAAACCCCCGGCACTCGTAAAACAGCAG 1170	(C-1/1

Fig. 2, continued

TOTAL STATEMENT OF THE ACTION OF THE ACCOUNT THE ACCOUNT CONTRACT	
GEGG THAGGGGG THE ALGERICATION COLOR OF THE ALGERICAL COLOR OF THE A	1 16-172
C T T C C T C T C T C T C T C C T C C T C C T C C C T C C C C C C C C C C C C C C C C C C C C	ec-ure
CONTICIOCO ABOCTACO ACCATOTO TIBATESCO I IACOCTIBACO ACCAGAGAGA CARABATIACO CARGADEA TICOTO CARGADEA CONTICATO CARGADEA CARGADO CARGADO CONTICATO CARGADO CARG	((C-L/1)
COMMISSION CACCIOCATE CONTROL TEMPORALICA CONTROL CONT	
COCCCOST TO CATEGO TRANSPORTAÇÃO MASTE TO TOTA DE TOTA	16-611
TIGIAEMICCCENCACEBOCICEMICACECIBECCICCICETRICCETRICCETRICEGERICAGECATEGCCATEGCAMENTITICE 12140 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 12240 1	EC-1/1
ITTISSECCECECC REGREGACE GENERAL RECOGNICISE PECCE STOCK SECRECACC SECREGACISE AND LEFON 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 12500 1	ee-113
ATERCTOCCOCCICCACSTGGAICACGAMATGGTTAGCAGATCCATGACCAGGACCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGGCAAGA	ec-115
CCCCCCCCECAGECCCCCCCCCCCCCCCCCCCCCCCCCC	IC-CF1
TICALECTIFF AND ICT CAMERAGE CHRACE INCICCICCOC (ACACAGOCET ACAIGE CEATE ACACAGOCET CATOCECCACC TICALECT TO 1270 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 12700 1	10-012
STATE (C-164) SECRETARIA DE CONTROL DE CONT	EC-AFA
CTETANET LE CETTE CETTE TE CAMMET MAGGE LA FEAMENT LE CETTE LA CETTE LA CETTE LA TETE CE ACTE LA CETTE	(C-184 (C-184
ATERICAL CONTROL OF THE PROPERTY OF THE PROPER	CC-8/4
Acchange Corta i Acces constitutes (constitutes constitutes consti	£C-814
SIALLIAN H	CC-111
PIE E E C E P PA LA PISPIS SO CRETICO	16-672
14 CECCACCTECCACACACTECTACACTECTACACTECTACACTECTACACTECTACACTECTACTACTACTACTACTACTACTACTACTACTACTACTAC	(C-1/1
STRACTURE CONTRACTURE CONTRACT	EC-172
ECCACETRACA CALACICITECCA TICAGA CALACET TO THE	ec-er-
which the first of	EC-4/4
TARREST TARREST AREA TO A FEW OLE O	£6.L/1

Fig. 2, continued

TRAGE CONTRECT TABLE TABLE TO THE TABLE TABLE THE TABLE	[C-\$f4
PASAY F F A Q Q S F Y R T C T S A S D A R R T S L C C L	EC-LF3
ATENCEATRICEMATECCACATECCATACCCCTETTYATCCEATGGAATGACGGCCATTTCFCCTCCTCTGGACACCCCTCTCGAGTATCCCC 19400 13700 13700 13720 13720 13720 13730 13740 13750 13750 13750 13751 TACTGCTACCGCTGTACCGCCCTCTATGGCGGAATAGGCTACTCTACCCCCCGCGTAAAAAAACACCACCATAGTACCACCACATACACCC TACTGCTACCGCTGTACCACCCCCTCTAAAAAAACACCACCACTAGTACACCCCCTCTAAAAAAAA	EC-LF2
IAGACATGANASICCAGACGCCTTCCS TOGGSTSTCISCCTCCGCCTTTGCCGTCCTCGTTBGCCACCACCACCACCACACGCCCTTT 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100 13100	EC-LFZ
CCATCETCETCACCCCCCACCACCACCETC AGGEARTETTEG TAFAAGCACAGGEBBGCCIGAGGCCCCCCCCACATCCCACCACCACCACGGGGA 13870 13870 13800 13800 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 13900 139	EC-CFE
CCIGGESTICTCEARACACTICGGAAIGASACGCAGCTACTCCTTGTGAGGCCTTTTTCAAAGGCCTTTATTAGGCTCTTAGSAAAGGCTTA 13900 13970 13980 13990 14090 14090 14010 14020 14030 14030 EGACCACABACCTTTGTGAACCTTTGTTGTGTCATGTGGAACATGTGGAAAGTCTTCCGAAATACTCTCCCGAAATACTCTTCCGAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCTTCCCAAATACTCCCAAATACTCCCAAATACTCTTCCCAAATACTCCCAAATACTCCCAAATACTCCCAAATACTCCCAAATACTCCCAAATACTCCAAATACTCCCAAATACTCCCAAATACTCCTTCCTCTTTTTT	£C-1\f2
TAGAAGASICCCCCAGACATGGITAAAACTGAGTCTCTGCCTCCCCAAGCAGTGCGGCGGGGGTTCTGGGATAGCGTCTTCTGC 14550 14550 14560 14570 14080 14070 14100 14100 14100 14100 14100 ATCTTCTCAGGGGGTCTGTTACCATTTTTTGGGTCAGACGGAGGGGGTTGTCAGCGCGCGC	{C-LF1
GTARROCTIGNAMACGETECTESCISCECISCECATECASETTOCTESCEAFGAGESTECGSCESCATELEGESTECETETTISETECT 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100 14100	16-171
EGITTIBETICATECASCECTANCE MATEGETTETTA FICITATETATETATA TORA TARGES ATTANTOCCI TECACCCICACCAS ASCEATETECT 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14200 14	EC-CF1
GARCAAGCICEICBATCCCGGGGTABACCATCIGCACCGCTGCTGCGGCAACTACCAGGATCA 14370 14370 14380 14380 14380 14380 14380 14380 14380 14380 14480 CTCGCTGGAGCAGCTAGGGGCCGTTGTGGAGCGGCCGAGCGGAGCGAGC	1C-LF1
CCACCCCCCCCCCABBACCAGARIGACCAIBCCGCCGAAGGGGTTTTIGAAGAAGAAGAAGAAACTGAAACCAGCGCCGCTAAACAAAC	(6-61)
CCCCCACCFICTEACTAGETIGATEATGACTAGCCACGCTACCCAGACTGTCCRTAAGTTCCCCCAGGCCGTCCACGAATTGATTTC 16500 12510 14500 16500 16500 16500 16500 16500 16500 16500 16500 400005TGCACGACTGATCCAACCTACTGACTGATCCAACAGCTTCAACAGCACTCCACCAGGCTCTAAAG 6 V	EC-LF3
TECCGITICACECTACATICECCAMATICE) ECCGLAGGE COCECCATORITETA COCCIDENA GITETA CICCECCACA TECCCICAL GET 1650 1650 1650 1650 1660 1660 1660 1660	(C-LF1
CAARRACETTERARECTACCTCCTCCCCCCCCCCCCCCCC	(C-LF1
TGAAGGICIECAGGETGGCAAFGCCGICCAGCTCGATGGTITTAAAGGGGGGTAGFCGTTGTAGACGIGGATCICGATTGCACTGGA 14796 14796 14790 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 14900 1	(C-1/1
AGTAGIACTGCCTGGTCGCCTGGCACACCTCCGTTATCTTTTTTGTGARGAABATCTCGTTFTCGGTGCCCAGTGTCTCCCTAGGTCT 14866 14870 14880 14890 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 14910 1	(C-LFS
THREE TOTAL TRANSCRIPT AND THE PROPERTY OF THE	CC-LFL
CONTRECEMENT ANCHEMACE ACTICES AS ACT CONTRECEMENT CONTRECEMENT ACCES (CONTRECEMENT ACCES (CONTRACEMENT AC	(c-tfi
TEACGATEGATTGAATTAATTITEATTAGTICTCTCAECACCATGTICTGCCTCTTCTCTCCCAGCACCAGGCCCCACAGGCCCCCCAAGGCCCCCCAAGGCCCCCC	fc-LF1
GEATECROTTEATCTEGECECECAGERATICETAGECCAAATTGEATCTBGACGETEGEGBAATTGTTEAGGGTGCCCAGGGACTTCCCGG 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 13200 1320	tc-L*1
CARCCATABAGEETACCACTATORTECECTOTCCCCACACTCCCACCACCACCACCACCACCACCACCACC	(C-LFt

Fig. 2, continued

ATGGGGCTGGGGGGTGGACGCCCCGGGGGGGGGGGGGGG	
ATGGGETTGEBEGGATEGACTECTEGEBEGTGEGAGTGEGAETGETAAGCTCTGTGAGETTGTTGAGGTGEGTGACGACGGGGGG TACCCGGCCCCCGTGAGGACGCCCCGCTTCTTCAGGTCTTCAGATTCGAGGTACTCCAGAGATTGCAGGTACTTCCACGCTCCTTCCT	EC-171
ACTICAL CALIFORNIA CONTROL STATE AND ACTICAL CONTROL C	EC-LF1
CGIACHICTEATGCATGGTCTIGTTCACCTGCTCTTICGATGCACTTGAAGGCGTCGGGAGGCTCTATGCCCACGGTTGTGTTGGTCACCA 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 15600 1560	fc-lf1
AGCIACHGE IGCCCICGICAGICACAAAIGIA IGACIICCCTOTTICIGICGGAIGCTCAAGICAAACIITICCCAGTOTIGAACC 1940 1940 1950 1970 1970 1970 1970 1970 1970 1970 197	£C-LF1
GECRETAGGLIGTCCTETTCICGAGCTICCAACATAGCETGIAAGTGCCCTIGTCCAGGAAGGCTCGGCGTTCGCCTITGCGGGTTCGTCC 15700 15700 15700 15900 15810 15800 15810 16800 16800 16800 16840 CCCCCCTCCCCCCACAGGAAGGAAGGAACGTCTAATCACCACATCAGCAGAAGGCTCCTCCCCCAACCGCAACGCGAACGCCCAACGCGAACGCCAACGCGAACGCCCAACGCGAACGCCCAACG	(- 1 / 1
CTECCTICETACTECTACTACTICACTACTICTCCCCCCACACACAGCICCCCCCCCCACACAGAAGGITICCCTATITTCCCCCCCAT 11810 15860 15810 15810 15810 15910 15910 15910 15910 15910 CACCCANACCATCACCCTCATACAACACTCATCAACACCACCACCAC	EG-(J1
AGAAAGGGACATTICCACAGICTGCCCGGGGGGGCGCCACAAGAGAGTGGAAGGGCCIGTTBBACTTGGCCATCATGTCAGTTATCAGGC 15400 15400 15400 15400 15400 15500 16500 16000 16000 16010 ICITICCCCTGTAAAGGGGCCACACCACGGGTGTTTCCTCCGGTCAGCCGGACAGCGGGACAGACCAGATCAATAGTCGG FF S M C T Q G C C C C C C C C C C C C C C C C C C	rc-tri
ACTIBACESTACTICTICTICTOTAACTCCATATCAACCACCOGGGCCETCATAGACCTCCGCTGCCCCCCGCGCGCGCCCCCCC 1830 1840 1840 1860 1860 1860 1800 1800 1800 1800 180	(C-LF1
TEECCETCCCCCCCCCTCCCCTTTAGETERACGTERATETTACCTCCCCGTCTACATACACCCCCCTCCTCACCCCATCTTTTGTCATCT 16170 16110 16140 16140 16140 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 16170 1617	tc.tr;
TEACCECEFTETACCACTECTACATEGIATCACT TEGTCABTT TEGTABCTOTCAACGEACAACTTC ICCTC ETECCCCT TEGTCACE 1610 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16270 16	
ACTICCECETACCACCCATTCTARATGAQAATRITGGTCACTATCTTGATCTAGAACCGCACCATTAACGAGTACGAATAATGFTGFCTT 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300 16300	EC-LF1
TARACACCATCARCAGECCC FECATBEGRIFETECECCARCGAGEGRACACTGGATGTCCCAGGGGAGAGEGRACACGAGAGTT 16300 16400 16400 16400 16400 16400 16400 1660 1650 16600 18630 18640 18670 ATTHEREFACE FIGURE CACACATERACAGGGGGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGG	{C-LF1
GACTGGGGARCTCCCAGACTEGGAAACGAAACTGGGTGGCGGACGGGGGGGGGG	CC-1.F1
CTRECRICTECOCACCARRACERCACGICARRACEGITAGCACCACCACCACCACCACCACGACCACCACCACCACCA	ec-tri
TECAGETICAE ANGECCATATAGATCCTTAAATACCCACCCCCTCACCCTTATACACAAAATATAGCCGGTCACATCACCTTAAATACCCACCC	
TAMASCICATCCBBBBBCACGGCCTBGCCGGGCCBACCGGCCCGCCGBCGCAGGATCAASEACCGCCCCCABETCTTBTTGTACAGG 18190 18180 18170 18180 18180 1880 1881 1880 18820 18818 1880 18818 ATTICGBETAGGCCCCCGTCCGGACCGGCCCGATGGCGCGGGGCGCGCGC	
EACACCETTACCACCECCCCCCCACACCCCCCCCACAAATTIGCTBCCTACATCACCCCCCCCCCCCCCCCCCCCCCCCCCC	
AACGMETICTATICAETCRETECETICATECTEGECATATCAGECAMETICAGGAATGCCTTCTGGTGGGGCCAGECTACCTGCGGGC 18730 18400 18400 18900 18900 17000 17010 TICTTCCAGAATGGGTCAGGGACCAGCTTGCTGGGCCGCTATAGGTCCCTTCAGGAAAACCAGCCCGTGGATGGA	
17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100 17100	
ACCESSION CONTROL OF THE PROPERTY OF THE PROPE	

Fig. 2, continued

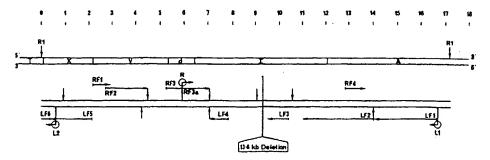


Figure 3. Organization of the EcoRI C fragment of B95-8 EBV. Open reading frames (RFI etc) are shown by horizontal thin arrows, AAUAAA sequences (vertical thin arrows) and the position of the deletion relative to EBV from Raji cells are also marked. The three promoters LI, L2 and R are indicated. The scale is in kb.

(2) Codon preferences

EBV has a G+C content of 58% (Weinberg & Becker, 1969; Schultze-Holthausen & zur Hausen, 1970) and the nucleotide composition of the EcoRI C fragment matches this closely (A 20.6%, G 29.1%, C 29.4%, T 20.9%). The pattern of codon usage in open reading frames detected in the DNA sequence can be used to predict whether these are actually translated. For any particular amino acid, the first two positions in its codons are nearly always fixed (except for the 6 codon families: leucine UUR, CUX; serine UCX, AGY and arginine CGX and AGR), so with a fairly random amino acid sequence the composition of the nucleotides in the first two positions will also be fairly random. In a coding region with a strong G+C bias the excess G+Cnucleotides will tend to be concentrated into the third position of the codons. So the nucleotide bias tends to be exaggerated in the codon usage and may produce a highly characteristic codon usage pattern. If the region is not being expressed into protein there is no such constraint pushing the excess G+C nucleotides into the third position and the codon usage pattern will be quite different. Such a highly biased codon usage pattern has been observed previously in bacteriophage ϕX with 31% T. In a random protein sequence encoded by a nucleotide sequence with a 58% G+C content, the G+C content of the first two codon positions should be 50% and the G+C content of the third position would be predicted to be 74%. Table 1 lists the % G+C content in the third position of the codons in all the reading frames shown in Figures 2 and 3. The average G+C content in the third position of the codons in these readings frames is 69%, approaching the third position G+C content calculated for a random protein sequence and substantially higher than the 58% of the whole sequence. This highly non-random distribution of the G+C content in the reading frames make it extremely likely that they or large elements of them correspond to protein coding regions. The biased codon usage will be detected by the FRAMESCAN program (Staden & McLachlan, 1982), which analyses the codon usage of a known gene or potential gene and uses the characteristic codon pattern to predict other similar possible coding regions. Thus we can quickly scan EBV sequences for similar putative coding regions

TABLE 1 Third position G+C in reading frames

Frame	Length	3rd G+C	% 3rd G+C
EC-RFI*	186	118	63-4
EC-RF2	571	445	77.9
EC-RF3	606	425	70·1
EC-RF3a	346	250	70·1 72·5
EC-RF4†	290	173	72·3 59·6
EC-LF1	858	651	75·9
EC-LF2	1016	832	
EC-LF3	313	229	81.9
EC-LF4	249	144	73·2
EC-LF5	608	369	57.8
EC-LF6t	116	71	60-7
		/1	61.2
verage excluding R	F3a, RF4:		69·1

Symbols: * only that region of the frame that does not overlap with EC-RF2 is included.

† See text, possibly spliced, but whole frame has been included here.

Only that part of the frame which is in the EcoRI C fragment is included.

and also use it to try to detect intervening sequences. This is particularly useful when the intervening sequence is "in frame" with the rest of the coding sequences and contains no termination codons. Below we describe the major features found in the DNA sequence and discuss the possible organization and expression of coding regions.

(3) Organization of reading frames

(a) Region 1-4500

The in vitro detected promoter L2 (Farrell et al., 1983) lies within the 3' end of the LF5 open reading frame (Figs 2 and 3). The first methionine codon preceded by a purine at -3 (Kozak, 1981) downstream of this promoter is found at the start of LF6 at position 350, one base beyond the termination codon at the end of LF5. The reading frame extends into the EcoRI H fragment and could code for a polypeptide of 717 amino acids (C. Séguin, personal communication). The EcoRI H fragment has been shown by sequence analysis to be contiguous with the C fragment. The expression of LF5 would depend on another leftward promoter. The absence of a nearby AATAAA sequence presumably would mean that it would be spliced somewhere at the 3' end. A possible splice donor sequence at position 390 has a seven out of nine base match with the consensus sequence of AGGTPuAGT.

On the R strand the methionine codon at the beginning of LF5 is overlapped with another methionine codon at the beginning of RFI. This frame remains open until position 2920 overlapping at its 3' end with another open frame RF2, which begins with the methionine codon at 2730. RF2 extends to the termination codon at 4440 and is followed by an AATAAA sequence 43 bases downstream. A possible promoter sequence for RF2 is situated upstream at position 2616 with the sequence

TATTTAAAAA and also the sequence GCCAATT at 2546. Interestingly the 22 base-pairs centred on position 2620 displays twofold symmetry with only two mismatched base-pairs. This potential promoter sequence was not detected by the *in vitro* transcription experiments of Farrell *et al.* (1983), although they did find some low level transcripts emanating from this region. Expression of RF2 could also be effected using the potential splice acceptor sequence two bases upstream of the first methionine codon. No obvious promoter-like sequences can be seen near the beginning of the RF1 and LF5 frames though potential splice acceptor sequences are found at position 2105 for RF1 and at 2244 for LF5. Both of these sequences occur (in the complementary sense) within the coding region of the other reading frame. Thus the possibilities exist for at least the independent expression of RF2 and possibly of a spliced RF1+RF2 frame. In the latter case a good potential splice donor sequence CAG/GTAAGC, which is an eight out of nine match with the donor consensus sequence, exists at position 2256 which is 28 codons downstream of the start of RF1.

(b) The region 4500-8000

(i) RF3 and 3a

The large open reading frame RF3 extends from the first available methionine codon at 5241 to the termination codon at 7056 which lies five codons past the AATAAA sequence at its 3' end. No in vitro promoter was detected upstream of the methionine codon although a potential promoter sequence TATTTAT occurs at 5035. Also possible splice acceptor sequences are found at positions 5059, 5131 and at 5218 before the methionine codon at 5241. There is, however, a promoter in the middle of the RF3 frame with a transcription start at approximately 5962-5972 (Farrell et al., 1983). The first methionine codon is found in frame at position 6021. Thus the potential exists to express the whole frame RF3 or the 3' half RF3a.

(ii) LF4

Twenty bases beyond the AATAAA sequence at the 3' end of RF3 is another AATAAA sequence on the L strand at position 7072. Associated with this is an open reading frame LF4 that could code for a protein of 248 amino acids starting at the methionine codon at 7839 and extending to the termination codon at 7095, 28 bases before the AATAAA sequence. The methionine codon is preceded by the sequence:

L strand 5'—CACACCCCACCACATATTTAGG-3'.

The pyrimidine-rich sequence 5' to the AG could serve as a splice acceptor sequence or the underlined sequence could correspond to the "TATA" promoter element. The preceding overlined sequence corresponds to the sequence found in the L2 and R promoters and in a number of other eukaryotic promoters (Farrell et al., 1983). The likely transcription start for this putative promoter would be centred on the 12 base twofold symmetric sequence 5' AGTTTTAAAACT 3' at position 7857. The amino acid sequence predicted from the LF4 reading frame contains many potential

glycosylation sites (Neuberger et al., 1972), in fact 12 out of the 14 asparagines in the sequence are followed two residues away with serine or threonine.

(iii) The B95-8 deletion region

The B95-8 strain of EBV is missing approximately 13.6 kb of DNA in the BamHI I fragment when compared with other strains (Raab-Traub et al., 1980). Part of the missing DNA has been shown to be homologous to a region of DNA in the BamHI H fragment, which is approximately 90 kb away. We have mapped the deletion point in the B95-8 sequence by also sequencing part of the Raji RI C fragment which does not carry the deletion. So far we only have sequence data for the 5' side of the deleted sequence so we do not yet know whether there are any repeated sequences involved in the deletion event. The data, however, allow us to define accurately the deletion point by comparison of the two sequences. The deletion is between residues 9326-7 of the B95-8 EcoRI C fragment. The Raji sequences will be published separately.

The region containing the deletion point between the LF3 and LF4 frames does not contain any open frames longer than 127 codons and we have not yet found any convincing pattern for its expression. An AATAAA sequence occurs on the R strand at 9081 and the largest open reading frame close to this is 64 codons. A possible promoter sequence CATAAAA occurs on the L strand at this point (9094). Other possible promoter sequences occur at 9544 (CATAAAA) and at 9574 (TATAATGA). A repetitive sequence occurs between 8550 and 8940 and the homologous sequences are shown in Table 2. A sequence of 14 consecutive Gs occurs in this region at 8799.

(c) The region 9470-17,172

(i) LFI

To the right of the deletion point most of the EcoRI C fragment sequence is taken up by three large open reading frames LF1-3. These are preceded by the L1 promoter where the transcription start has been mapped to approximately 16,650. The first methionine codon preceded by a purine at -3 occurs at 16,636 at the beginning of the LF1 frame. This frame ends 41 bases before the AATAAA sequence at 14021. However, using the program FRAMESCAN (Staden & McLachlan, 1982) the codon usage pattern changes considerably between 15,550 and 15,490, and to a lesser extent to 15,374, even when using the LF1 coden pattern itself as the input reference sequence. This region contains on the R strand the canonical promoter sequences GGCCAAC (15,464) and TATAAAATAT at (15,518) as noted by Farrell et al. (1983) but not detected as an in vitro promoter by them. This may indicate that this region is spliced out of the LF1 frame and lends some support to the idea that this region is an in vivo rightward promoter that is not detected by the in vitro system. There is on the L strand a possible splice donor sequence at 15,601 and possible acceptor sequences at 15,414 and 15,257 that would encompass the potential promoter region. If this region does function as a promoter it would presumably splice to the reading frames downstream at the end of the EcoRI C fragment and in the D het fragment in order to

TABLE 2

Repeated sequence elements in region 8550-8912 of EcoRI C

avoid the in phase initiation and termination codons in the 150 bases downstream of the possible transcription initiation point.

(ii) LF2

LF2 extends from the methionine codon (14,060) two bases after the LF2 termination codon, past the LF1 AATAAA sequence to the termination codon at 11,015. No candidate for a promoter for LF2 can be predicted so that expression of this frame may be effected by splicing from a leftward promoter such as L1. A possible splice acceptor sequence occurs at 14,137. No AATAAA sequence is found at the 3' end of LF2 so it would presumably be spliced at this end as well. A 22 base-pair sequence with perfect twofold symmetry is found at positions 10,951–10,972 at the end of the LF2 frame.

(iii) LF3

This frame lies approximately 600 bases beyond the 3' end of LF2 starting at the methionine at 10,413 and extending to 9475 about 150 bases before the deletion point. No promoter sequences are found but a possible splice acceptor sequence is situated at 10,444. There are two possible initiation codons at the start of the frame at 10,413 and 10,401. No AATAAA sequence is present near the 3' end although this is close to the deletion point. A possible splice donor sequence occurs approximately 50 bases before the end of the frame at 9526. A membrane role is predicted from the amino acid sequence which contains a regular array of hydrophobic regions separated by regions of charged amino acids (see below).

(iv) RF4

An open reading frame overlapping LF2 is found on the R strand. No promoter or AATAAA sequences are found near this frame. Several possible donor and acceptor splice sequences are present (13,038, 13,163, 13,190, 13,320 and 13,403) suggesting that perhaps only parts of the overlapping reading frame RF4 might be used.

Although the sequence of the EcoRI D het fragment is known (unpublished results) we have not yet overlapped the C and D het sequences to exclude the possibility of a small undetected EcoRI fragment between them. Therefore we have not correlated the D het reading frames with those occurring between the L1 promoter and the 3' end of the EcoRI C fragment.

(4) AATAAA sequences

RF2, RF3, LF1 and LF4 all have AATAAA sequences close to their 3' ends suggesting that these sequences play a role in their mRNA maturation and polyadenylation. Several other AATAAA sequences are found in the *EcoRI* fragment; it is difficult to predict whether these are functional because AATAAA sequences are sometimes found in coding and non-coding sequences of mRNAs. Benoist *et al.* (1980) noted that there was a homologous sequence in the 30 bases downstream of the

AATAAA sequence in some mRNAs with the consensus sequence TTTTCACTGC. Except for a seven out of ten match with a sequence near the AATAAA at 10,573 neither those AATAAA sequences associated with the ends of reading frames nor the isolated AATAAA sequences displayed any homology with this consensus sequence. In some cases possible splice acceptor sequences are seen associated with short reading frames near the other AATAAA sequences. Three such sequences occur upstream from the AATAAA (position 922) at 792, 840 and 877 and in another case at 9049 preceding the AATAAA sequence at 9081. Exceptions to the AATAAA sequences are known; several cases of ATTAAA (listed by Ahmed et al., 1982) and one of ATAA (Setzer et al., 1980). No ATTAAA sequences are found in the EcoRI C fragment and we have not searched for the sequence ATAA on the grounds that it is functionally very rare and not unique enough, i.e. not long enough to consider all possible occurrences in this sequence.

(5) Coding capacity

The EcoRI C fragment represents approximately 10% of the EBV genome and the sequence analysis of this fragment provides the first opportunity to assess the coding potential and arrangement of the genome. From an analysis of the sequence we would predict at least six to eight proteins coded in this region. However, with multiple choice splicing as found in adenovirus this could be a conservative estimate. The open reading frames cover most of the DNA sequence and their arrangement with respect to each other coupled with their similarity of codon usage (enhanced third position GC) lead us to expect that most if not all of the indicated open frames are expressed into protein. The conclusion is further strengthened by the finding of AATAAA sequences near the 3' ends of half of the large open reading frames which could participate in their mRNA maturation and polyadenylation. The lack of these sequences at the ends of the other open frames coupled with the lack of obvious promoter sequences near the 5' end of most of the frames would suggest that RNA splicing is involved in their putative expression. In most cases there are good candidates for splice acceptor and donor sites near their 5' and 3' ends, respectively. However, given the possible redundancy in these junction sequences, these are difficult to predict with any certainty.

There are several examples of coding economy in the sequence, which is perhaps surprising in such a large viral genome. The termination codon for LF1 and the possible initiation codon for LF2 are only two bases apart. Similarly those for LF5 and LF6 are only one base apart. The possible initiation codons for RF1 and LF5 are overlapping on the opposite strands. The AATAAA sequences associated with the frames RF3, 3a and LF4 which face each other on opposite strands are only 20 bases apart. Thus the mRNA ends would be expected to overlap on the opposite strands by about 20 bases. Also the closeness of the AATAAA sequences to the termination codons are characteristic of the compactness of viral genomes.

(6) Transcription

Three promoters were mapped by Farrell et al. (1983) in this fragment using the HeLa cell extract described by Manley et al. (1980). This in vitro system has been

shown to be an accurate reflection of the *in vivo* situation for some but not all promoters. Thus, for example, late viral promoters depending on a viral function might not be expected to be active in this system. This situation has been argued for another herpes virus HSV1 gene (Frink et al., 1981). Thus it is possible that this region contains other promoters not detected *in vitro*. A number of possibilities have been mentioned previously in the text. If a viral product is indeed required to turn on transcription from these promoters then their detection may depend on the development of an EBV *in vitro* system using a productively EBV infected cell extract.

Recently, Hummel & Kieff (1982a) have mapped a large number of early and late B95-8 RNA transcripts by hybridization to BamHI and EcoRI restriction enzyme fragments. In the EcoRI C fragment there are five BamHI sites so we can attempt to correlate some of the transcripts with the open reading frames. In BamHI T and X (W in notation of Hummel & Kieff) a 5.0 kb early and a 2.7 kb late transcript have been mapped. These do not extend into the BamHI V fragment (see Fig. 2) (V1 in their nomenclature) so that it is likely that these transcripts come from the LF5 and LF6 reading frames. RF1 and RF2 are almost totally situated in BamHI V (V1). Four transcripts are mapped to this fragment of sizes 6.4 kb, 5.5 kb, 3.2 kb and 2.5 kb. As this fragment is only approximately 3 kb in length the relationship of the four transcripts to the RF1 and RF2 frames is unclear. Also these transcripts do not seem to hybridize to the adjacent BamHI fragments, so they could represent multiple splicing events from outside this region to the RF1 and RF2 reading frames. The BamHI d (b) fragment hybridizes to a 1.4 kb late transcript and it is likely that this represents the RF3a reading frame transcribed from the R promoter. BamHI I contains two transcripts, the 1.4 kb late transcript assigned to RF3a and a 1.1 kb late transcript. LF4 and LF3 and part of LF2 are in this fragment. LF3 and LF4 are suitable sizes with LF4 being the best candidate with a possible promoter or splice acceptor sequence at the 5' end and a AATAAA sequence at the 3' end. No other transcripts were detected that could correspond to the large LF2 frame. A 3-0 kb RNA has been mapped to the BamHI A fragment of the EcoRI C fragment. This could correspond to the LFI (857 codons) or the LF2 (1015 codons) frames.

Therefore, although some correspondence can be found between the two sets of mapping data quite a few anomalies remain. It should now be possible to establish an accurate transcription map from this region by a combination of probing the fractionated RNAs with the substantial bacteriophage M13 clone bank of over 500 clones (each containing approximately 400 bases of EcoRI C DNA) that were used to generate the EcoRI C DNA sequence and by sequencing a cDNA library of the viral transcripts.

(7) Early membrane antigens and the B95-8 deletion

B95-8 has two unique features not found in other EBV strains, the 13-6 kb deletion and different proportions of the membrane antigens gp350 and gp220 (which are antigenically related). There is considerable interest in these membrane antigens because antibodies to them can neutralize EBV infections in tissue culture (Hoffman et al., 1980; Thorley-Lawson & Geilinger, 1980; Thorley-Lawson & Poodry, 1982). These glycoproteins or parts of them might therefore be used for immunizing people against

EBV infection. Such a procedure might be applied to EBV seronegative adolescents to reduce the risk of contraction of infectious mononucleosis. Another use of such an immunization might be to assess the causative role of EBV infection in African Burkitt's lymphoma by immunizing very young children from the appropriate geographic region against EBV infection and comparing the subsequent incidence of the disease among them with a non-immunized group (Epstein, 1976). The membrane antigens gp350 and gp220, which are designated VE1 and VE2 by Wells et al. (1982), are also thought to play an important role in the extreme selectivity of EBV binding to cell surfaces.

Whether there is a direct genotype/phenotype relationship between the DNA deletion and altered expression of the membrane antigens is uncertain, but the two leftward reading frames (LF3 and LF4) around the deletion point have features consistent with a membrane glycoprotein. Computerized analysis of the hydropathy (Kyte & Doolittle, 1982) shows that LF3 has alternating stretches of hydrophobic amino acids with charged residues in between and is probably a membrane protein. The most likely arrangement for this polypeptide in a membrane is shown in Figure 4. LF4 has a large number of potential glycosylation sites (sequence Asn-X-Thr/Ser). It is possible that some or all of the leftward reading frames could be spliced together to produce the antigenically related glycoproteins of molecular weights 350,000 and 220,000 (gp350 and gp220). The presence of the deletion between LF3 and LF4 could in some way affect the levels of expression of the reading frames in this region.

The fact that the intact glycoproteins contain carbohydrate means that we cannot directly correlate the sizes of the polypeptides predicted from open reading frames with the sizes of the proteins. A similar problem arises in trying to correlate proteins produced by in vitro translation (where little or no glycosylation occurs) with the

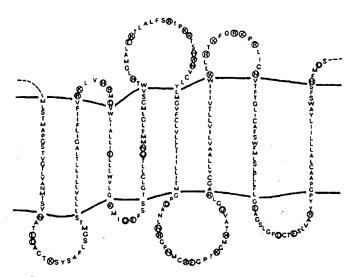


Figure 4. Putative folding of the LF3 polypeptide through a lipid bilayer.

complete proteins found in vivo. Perhaps by cloning sections of the viral DNA into eukaryotic expression vectors, transfecting those into suitable tissue-culture cells and looking for production of the complete viral protein it will be possible to assign proteins unequivocally to the DNA sequence.

Acknowledgments

During this work P.J.F. held a Medical Research Council training fellowship and P. D. was a fellow of the North Atlantic Treaty Organization.

References

- Ahmed, C. M. I., Chanda, R. S., Stow, N. D. & Zain, B. S. (1982). The nucleotide sequence of mRNA for the M_r 19,000 glycoprotein from early gene block III of adenovirus 2. Gene 20, 339-346.
- Arrand, J. R., Rymo, L., Walsh, J. E., Bjork, E., Lindahl, T. & Griffin, B. E. (1981). Molecular cloning of the Epstein-Barr virus genome as a set of overlapping restriction endonuclease fragments. Nucl. Acids Res. 9, 2999-3014.
- Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980). The ovalbumin gene-sequence of putative control regions. Nucl. Acids Res. 8, 127-142.
- Birnboim, H. C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. 7, 1513-1523.
- Bornkamm, G. W., Deluis, H., Zimber, V., Hudewentz, J. & Epstein, M. A. (1980). Comparison of Epstein-Barr virus strains of different origins by analysis of the viral DNAs. J. Virol. 35, 603-618.
- Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P. & Kedinger, L. (1980). Promoter sequences of eukaryotic protein-coding genes. Science 209, 1406-1414.
- Dambaugh, T., Beisel, C., Hummel, M., King, W., Fennewald, S., Cheung, A., Heller, M., Raab-Traub, N. & Kieff, E. (1980). Epstein-Barr virus (B95-8) DNA VII: molecular cloning and detailed mapping. *Proc. Nat. Acad. Sci., U.S.A.* 77, 2999-3003.
- Deininger, P. L. (1983). Random subcloning of sonicated DNA; application to shotgun DNA sequencing. Anal. Biochem. 129, 216–223.
- Delius, H. & Bornkamm, G. W. (1978). Heterogeneity of Epstein-Barr Virus III. Comparison of a transforming and a nontransforming virus by partial denaturation mapping of their DNAs. J. Virol. 27, 81-89.
- Duckworth, M. L., Gait, M. J., Goelet, P., Hong, G. F., Singh, M. & Titmas, R. C. (1981). Rapid synthesis of oligooxyribonucleotides VI. Efficient, mechanised synthesis of heptadecadeoxyribonucleotides by an improved solid phase phosphotriester route. Nucl. Acids Res. 9, 1691-1706.
- Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C. & Proudfoot, N. J. (1980). Structure and evolution of the human β-globin gene family. Cell 21, 653-668.
- Epstein, M. A. (1976). Implications of a vaccine for the prevention of Epstein-Barr virus infection: ethical and logistic considerations. Cancer Res. 36, 711-714.
- Epstein, M. A. & Achong, B. G. (1979). Editors of The Epstein-Barr Virus, Springer-Verlag, Berlin.
- Farrell, P. J., Deininger, P. L., Bankier, A. & Barrell, B. (1983). Homologous upstream sequences near Epstein-Barr virus promoters. *Proc. Nat. Acad. Sci., U.S.A.*, 80, 1565-1569.
- Frink, R. J., Draper, K. G. & Wagner, E. K. (1981). Uninfected cell polymerase efficiently transcribes early but not late herpes simplex virus type I mRNA. Proc. Nat. Acad. Sci., U.S.A. 78, 6139-6143.

110-

Hayward, S. D. & Kieff, E. D. (1977). DNA of Epstein-Barr Virus. II. Comparison of the molecular weights of restriction endonuclease fragments of the DNA of Epstein-Barr virus strains and identification of end fragments of the B95-8 strain. J. Virol. 23, 421-429.

Hayward, S. D., Nogee, L. & Hayward, G. (1980). Organization of repeated regions within the Epstein-Barr virus DNA molecule. J. Virol. 33, 507-521.

Heller, M., Dambaugh, T. & Kieff, E. (1981). Epstein-Barr Virus DNA IX. Variation among viral DNAs from producer and non-producer infected cells. J. Virol. 38, 632-648.

Hoffman, G. J., Lazarowitz, S. G. & Hayward, S. D. (1980). Monoclonal antibody against a 250,000-dalton glycoprotein of Epstein-Barr virus identifies a membrane antigen and a neutralizing antigen. Proc. Nat. Acad. Sci., U.S.A. 77, 2979-2983.

Hummel, M. & Kieff, E. (1982a). Epstein-Barr Virus RNA VIII. Viral RNA in permissively infected B95-8 cells. J. Virol. 43, 262-272.

Hummel, M. & Kieff, E. (1982b). Mapping of the polypeptides encoded by the Epstein-Barr Virus genome in a productive infection. *Proc. Nat. Acad. Sci., U.S.A.* 79, 5698-5702.

Kozak, M. (1981). Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. Nucl. Acids Res. 9, 5233-5252.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105-132.

Manley, J. L., Fire, A., Cano, A., Sharp, P. & Gefter, M. (1980). DNA-dependent transcription of adenovirus genes in a soluble whole-cell extract. Proc. Nat. Acad. Sci., U.S.A. 77, 3855-3859.

Messing, J. & Vieira, J. (1982). A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene 19, 269-276.

Messing, J., Crea, R. & Seeburg, P. H. (1981). A system for shotgun DNA sequencing. Nucl. Acids Res. 9, 309-321.

Miller, G., Shope, T., Lisco, H., Stitt, D. & Lipman, M. (1972). Epstein-Barr virus: transformation, cytopathic changes and viral antigens in squirrel monkey and marmoset leukocytes. Proc. Nat. Acad. Sci., U.S.A. 69, 383-387.

Mount, S. M. (1982). A catalogue of splice junction sequences. Nucl. Acids Res. 10, 459-472.
 Neuberger, A., Gottschalk, A., Marshall, R. D. & Spiro, R. G. (1972). In The Glycoproteins: Their Composition, Structure and Function (Gottschalk, A., ed.), Elsevier, Amsterdam, pp. 450-490.

Pritchett, R. F., Hayward, S. D. & Kieff, E. D. (1975). DNA of Epstein-Barr Virus I. Comparison of DNA of virus purified from HR-1 and B95-8 cells: size, structure and relatedness. J. Virol. 15, 556-569.

Raab-Traub, N., Pritchett, R. & Kieff, E. (1978). DNA of Epstein-Barr virus III. Identification of restriction enzyme fragments that contain sequences which differ among strains of Epstein-Barr virus. J. Virol. 27, 388-398.

Raab-Traub, N., Dambaugh, T. & Kieff, E. (1980). DNA of Epstein-Barr virus VIII: B95-8, the previous prototype, is an unusual deletion derivative. Cell 22, 257-267.

Sanger, F. & Coulson, A. R. (1978). The use of thin acrylamide gels for DNA sequencing. FEBS Letters 87, 107-110.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain terminating inhibitors. Proc. Nat. Acad. Sci., U.S.A. 74, 5463-5468.

Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. (1980). Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 143, 161-178.
Schultz Helthauser, H. & Strander, M. (1970). P. Schultz Helthauser, H. (1970). P. Schultz Helthauser, H. (1970).

Schulte-Holthausen, H. & zur Hausen, H. (1970). Partial purification of the Epstein-Barr virus and some properties of its DNA. Virology 40, 776-779.

Setzer, D. R., McGrogan, M., Nunberg, J. H. & Schimke, R. T. (1980). Size heterogeneity in the 3' end of dihydrofolate reductase mRNAs in mouse cells. Cell 22, 361-370.

Shenk, T. (1981). Transcriptional control regions; nucleotide requirements for initiation by RNA polymerases II and III. Current Topics Microbiol. Immunol. 93, 25-46.

Skare, J. & Strominger, J. L. (1980). Cloning and mapping of BamHI endonuclease fragments of DNA from the transforming B95-8 strain of Epstein-Barr virus. Proc. Nat. Acad. Sci., U.S.A. 77, 3860-3864 and 7510.

- Staden, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucl. Acids Res. 10, 4731-4751.
- Staden, R. & McLachlan, A. D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucl. Acids Res. 10, 141-156.
- Sugden, B., Summers, W. C. & Klein, G., (1976). Nucleic acid renaturation and restriction endonuclease cleavage analyses show that the DNAs of a transforming and a nontransforming strain of Epstein-Barr virus share approximately 90% of their nucleotide sequences. J. Virol. 18, 765-775.
- Thorley-Lawson, D. A. & Geilinger, K. (1980). Monoclonal antibodies against the major glycoproteins (gp 350/220) of Epstein-Barr virus neutralise infectivity. Proc. Nat. Acad. Sci., U.S.A. 77, 5307-5311.
- Thorley-Lawson, D. A. & Poodry, C. A. (1982). Identification and isolation of the main component (gp350-gp220) of Epstein-Barr virus responsible for generating neutralising antibodies in vivo. J. Virol. 43, 730-736.
- Tooze, J. (1980). Editor of *DNA Tumor Viruses*. Cold Spring Harbor Publications, New York. Weinberg, A. & Becker, Y. (1969). Studies on EB virus of Burkitt's Lymphoblasts. *Virology* 39, 312-321.
- Wells, A., Kiode, N. & Klein, G. (1982). Two large virion envelope glycoproteins mediate Epstein-Barr virus binding to receptor-positive cells. J. Virol. 41, 286-297.
- Wieslander, L. (1979). A simple method to recover intact high molecular weight RNA and DNA after electrophoretic separation in low gelling temperature agarose gels. *Anal. Biochem.* 98, 305-309.
- zur Hausen, H., Hecker, E., O'Neill, F. J. & Freese, U. K. (1978). Persistent oncogenic herpes virus induced by the tumour promoter TPA. Nature (London) 272, 373-375.
- zur Hausen, H., Bornkamm, G. W., Schmidt, R. & Hecker, E. (1979). Tumor initiators and promoters in the induction of Epstein-Barr virus. *Proc. Nat. Acad. Sci., U.S.A.* 76, 782-785.

THIS PAGE BLANK (USPTO)